Knowledge organisation and information retrieval using Galois lattices

Laszlo Szathmary, Amedeo Napoli

Knowledge organisation and information retrieval using Galois lattices

Laszlo Szathmary and **Amedeo Napoli**[1]

Abstract. In this paper[2] we investigate the application of Galois (or concept) lattices on different data sources (e.g. web documents or bibliographical items) in order to organise knowledge that can be extracted from the data. This knowledge organisation can serve a number of purposes (e.g. knowledge management in an organisation, document retrieval on the Web, etc.). Galois lattices can be considered as classification tools for knowledge units in concept hierarchies that can be used within a knowledge-based system. Moreover, Galois lattices can be used in parallel with domain ontologies for building more precise and more concise concept ontologies, and for guiding the knowledge discovery process.

1 Introduction

Today knowledge management (KM) has become one of the key progress factors in enterprises[3]. We have made experiments on KM within our research team, because it can also be considered a small enterprise. Experience frequently shows that researchers, even within the same team, do not know exactly what other people are working on. We aim to analyse the global work of the team to find interconnections between members, to know which are the main/marginal works in the team, i.e. to carry out a diagnosis on the research work.

In a research team, publications are the best way to describe the interests of a person. This is why we have chosen to analyse the bibliographical items. We have worked with the BibTeX descriptions that provide us metadata about a paper, e.g. title, authors, keywords, abstract, year of publication, etc. A BibTeX entry is similar to Dublin Core. The Dublin Core metadata standard is a simple yet effective element set for describing a wide range of networked resources, especially HTML pages [4]. BibTeX descriptions are a standard for scientific papers, and having a "controlled vocabulary", which is a limited set of consistently used and carefully defined terms, they can be interpreted in terms of the Dublin Core.

To analyse the publications, we have used classification as data mining technique in the knowledge discovery process. For classification we have used Galois lattices connected with domain ontologies. Generally, ontologies provide a shared and common understanding of a domain. In our case, we have built some ontologies to explain our knowledge about the team's members and publications. With ontologies, a more intelligent way of managing knowledge and searching documents, based on the content (semantics) of the manipulated document, can be performed.

1 LORIA/Inria-Lorraine, 615 rue du Jardin Botanique, BP 101, F-54600, Vandœuvre-lès-Nancy, France, {szathmar,napoli}@loria.fr

2 This publication was made in the frame of the French-Hungarian research program Balaton (Balaton F-23/03).

3 We will use the term "enterprise" to denote a set of people working together and having a common goal, e.g. machine design, sales, research, ...

In this paper we introduce a method joining Galois lattices and ontologies to guide the knowledge discovery process between people, documents and topics. We propose to use domain ontologies for data cleaning and for data mining connected with concept lattices. Moreover, we present a multidimensional basis for building lattices.

The rest of the paper is organized as follows. In Section 2 we give an overview of ontologies and we describe our domain ontology. Section 3 defines knowledge discovery, presents the role of ontologies in the KDD (knowledge discovery in databases) process, and gives an introduction to the basics of concept lattices. In Section 4 we discuss our new approach of data mining with Galois lattices and ontologies. In this section we introduce our multidimensional basis for building lattices, and we give some examples. In Section 5 we provide a synthesis and a list of related works. Section 6 draws conclusion and addresses some perspectives for the future research.

2 Ontologies

Ontologies were developed in artificial intelligence to facilitate knowledge sharing and reuse [21]. An ontology is a shared and common understanding of some domain that can be communicated among people and application systems. An ontology is an explicit "specification of a shared conceptualization" [12]. *Ontologies represent knowledge about domains*. They 1) identify the key concepts in a domain, 2) identify key relations between these concepts, and 3) identify a vocabulary for the concepts and relations. They specify the meaning of the vocabulary terms precisely enough so that they can be shared between a) humans and humans, b) humans and machines, and c) machines and machines.

To explain and access knowledge about our team we needed some ontologies. Several simple ontologies are freely available on the Web, like the DAML Ontology Library[4] or the DMOZ [5] (Directory Mozilla) effort to generate large simple ontologies. Since an appropriate ontology of scientific keywords covering our research area was not available, we have decided to build one. Our initial idea was to reuse some existing ontologies by merging them with an ontology mapping method [16].

An extract of our ontology is illustrated in Figure 1. A node is a concept, and a directed edge represents the class–subClassOf relation. All concepts are subclasses of a root concept, called "Top". Keywords of our publications are stored in the concepts. For more details see Section 4.1.

4 `http://www.daml.org/ontologies/`

5 `http://www.dmoz.org`

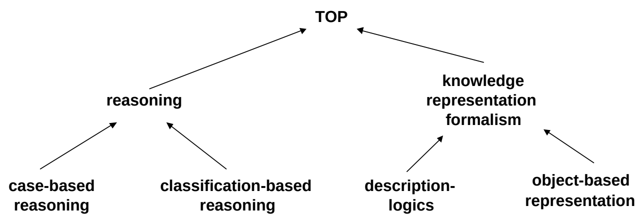

Figure 1. Ontology of keywords

3 Knowledge discovery in databases

In this section we define the terms "knowledge discovery" and "data mining". Then, we list what an ontology can be used for in the KDD process.

3.1 Definition of KDD

Knowledge discovery and data mining are techniques to discover strategic information hidden in (very large) databases. The terms *knowledge discovery in databases* and *data mining* are often used as synonyms. Actually KDD is a process for finding valid, useful and understandable patterns in data. Data mining is just a part of this process used for the extraction of patterns from data [11]. Additional steps include data preparation, data selection, data cleaning, and interpretation/evaluation of the results to derive knowledge at the end [5].

3.2 Ontologies in the KDD process

Using ontologies within the KDD process we can ensure:

- Data cleaning. It allows mapping data to a single naming convention, and handling noise and errors when possible (see Section 4.1).
- Knowledge organisation (with Galois lattices). Formal concept analysis is a classification method in data mining. After having built a Galois lattice, automatic extraction of rules between the attributes of objects is possible (see Section 4.2.2 and Section 4.2.3).
- Information retrieval. Query answering can be carried out on bibliographical items, for instance ranking them by their relevance on a keyword [3].
- Creating a family of lattices based on the depth of properties in the ontology. Different lattices can be constructed by descending in the ontology from the Top concept and taking into account more and more concepts.

In this paper we only focus on the first two items in this list.

3.3 Galois lattices

Galois (or concept) lattices provide a natural and formal setting to discover and represent concept hierarchies. "Formal concept analysis" is mainly used for data analysis, i.e. for investigating and processing explicitly given information [8, 7, 9]. As input data, we consider a binary relation between a set of individuals and a set of properties, e.g. a set of documents (D) and a set of keywords (W). A *context* is a triple (D, W, R), where $R \subseteq D \times W$. $R(d, w)$ means: the document d has the keyword w. We can think of the set of keywords associated with a document as a bit vector where each bit is on or off depending on whether a document has the keyword or not.

From such a binary correspondance one may derive for each document its keyword pattern (that is, the set of all keywords that apply to it). Similarly, one may derive for each keyword its document pattern (that is, the set of all documents to which it applies) [13].

A *concept* (C) is determined by its extension and intension: $(Extension(C), Intension(C))$, where the extension consists of all objects that share the attributes in the intension, and vice versa, the intension consists of all attributes that are common to the objects in the extension. These concepts are *formal*, which means that they are mathematical entities and must not be identified with concepts of the mind.

The *subsumption relation*, or also called *partial ordering* ($\leq$) is defined between concepts: $C_1 \leq C_2$ (C_1 is subsumed by C_2, or C_2 subsumes C_1), iff: $Extension(C_1) \subseteq Extension(C_2)$, and dually $Intension(C_2) \subseteq Intension(C_1)$. A *Galois lattice* is a set of concepts defined by the context (D, W, R), and organised by the subsumption relation ($\leq$).

In the context (D, W, R) let X and $Y \subset W$, where $X \cap Y = \emptyset$. An association rule is an implication of the form $X \Rightarrow Y$. The rule $X \Rightarrow Y$ is true in the context (D, W, R) iff every document in D that contains the keywords X also contains the keywords Y. For further details on association rules see [1, 17, 22].

Example: let us consider a context (D, W, R), where $D = \{d_1, d_2, d_3\}$ and $W = \{w_1, w_2, w_3, w_4\}$. In the following table *1* means that a document contains the given keyword (see Table 1).

	w_1	w_2	w_3	w_4
d_1	1		1	
d_2	1	1		
d_3			1	1

Table 1. A simple example

The documents d_1 and d_3 are the only documents that have the keyword w_3, and vice versa, the keyword w_3 is possessed only by documents d_1 and d_3. Thus $\{d_1, d_3\} \times \{w_3\}$ is a concept of the Galois lattice. Figure 2 shows the so-called Hasse diagram of the Galois lattice. In this diagram each node is a concept and an edge represents the subsumption relation ($\leq$). Two concepts are called *neighbors* if there is a direct edge between them in the Hasse diagram. We can also determine two association rules in this lattice: $w_2 \Rightarrow w_1$, and $w_4 \Rightarrow w_3$.

We can see that a Galois lattice has two special concepts. The *Top* concept ($\{d_1, d_2, d_3\} \times \{\emptyset\}$) has an extension that contains all the objects, and the *Bottom* concept ($\{\emptyset\} \times \{w_1, w_2, w_3, w_4\}$) has an intension that contains all the attributes. To construct the lattice we have used the incremental algorithm described in [10]. It builds a lattice by adding a new object to an already existing Galois lattice, without reconstructing the lattice from scratch.

4 Galois lattices connected with ontologies

4.1 Data cleaning with ontologies

In this section, we detail why and how an ontology can be used for enhancing a keyword-based information retrieval, just as a thesaurus

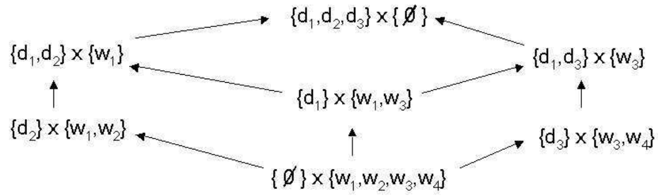

Figure 2. Hasse diagram of the lattice

does. Using just a simple keyword-based search over the publications may lead to several problems:

- Based on an exact string matching, the documents that are returned must be indexed by the exact keyword used for the search. Moreover, if a syntactically incorrect keyword is attached to a document (typo) then even if a correct search term is used, the appropriate document cannot be found.
- Synonyms: keywords may have various synonyms, and, as the association of a keyword to a document is not based on definite rules or a grammar, more than one keyword may be attached to a document for the same topic.
- Languages: as we are working with a bibliography where there are at least two languages, French and English, keywords may be used in both languages. This fact must also be taken into account within the search.

Using an ontology we can solve these problems by grouping the same/similar keywords in a concept (Figure 3). Now if we perform a search on “DL” for instance, it will give a much more precise result containing documents with all these keywords.

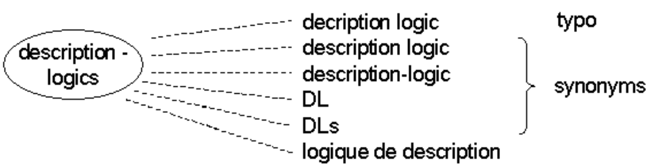

Figure 3. Grouping keywords

We used the ontology for data cleaning in the first step of the knowledge discovery process: all the keywords that were present in the publications are filtered through the ontology and then a reduced keyword set is obtained (Figure 4). For the construction of Galois lattices we took this keyword set. Grouping keywords as shown in Figure 3 can solve the three problems mentioned above. However, extending this kind of search to the Web is not really satisfactory yet. We can still filter the keywords that are recognized by the ontology and keep the rest of the keywords, but the list of keywords in the ontology cannot be complete and documents can be missed. In this case an approximate or inexact matching may be mandatory.

Until now, we have analysed 147 publications. These publications had altogether 335 different keywords that we managed to insert in an ontology having 89 concepts. Then the keyword set is reduced to 89 different keywords. For instance, taking the following original keywords (‘DL’, ‘DLs’, ‘case-based problem solving’, ‘CBR’, ‘galois connection’) a filtering through the ontology produces the following result: ‘description logics’, ‘case-based reasoning’, ‘Galois lattices’.

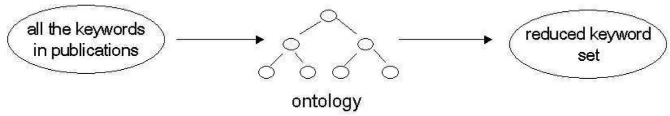

Figure 4. Filtering keywords

4.2 Document organisation based on Galois lattices

In this subsection we present a multidimensional basis for building Galois lattices. This basis allows the construction of different lattices. We give examples for some of these lattices.

4.2.1 A multidimensional basis for building lattices

“Formal concept analysis” is mainly used for data analysis, i.e. for investigating and processing explicitly given information [8, 7, 9]. Generally, we have a binary relation $R(x, y)$, where $x \in X$ and $y \in Y$. We are looking for interactions between x and y, and for this we study the relation between x and y. Galois lattices can discover relations, since they permit to find the “classes of x” that are in relation with the “classes of y”.

In our team we have studied the relations between individuals (team members, X), publications (documents, Y) and keywords (Z) to find answers to the following questions:

1. Which persons have published together? Which are these documents?
   For answering this question, the relation R_1 between individuals and publications has to be studied: $R_1(x \times y)$, where $x \in X(individuals)$ and $y \in Y(documents)$.
2. Which persons work on a common topic? What is this topic?
   For answering this question, the relation R_2 between individuals and keywords has to be studied: $R_2(x \times z)$, where $x \in X(individuals)$ and $z \in Z(keywords)$.
3. Which documents are written about a common topic? What is this topic?
   For answering this question, the relation R_3 between publications and keywords has to be studied: $R_3(y \times z)$, where $y \in Y(documents)$ and $z \in Z(keywords)$.

In this article we will detail the second and third point.

People, documents and keywords can also be considered as axes in a multidimensional system (see Figure 5). In this system different lattices can be constructed on the planes defined by the axes, depending on the relation to be analysed.

Another way of regarding this knowledge organisation problem is to consider three-dimensional points in this multidimensional system. A point (x_0, y_0, z_0) exists if and only if the following is true: $R_1(x_0, y_0) \wedge R_2(x_0, z_0) \wedge R_3(y_0, z_0)$. The binary relations R_1, R_2 and R_3 are supposed to be symmetric.

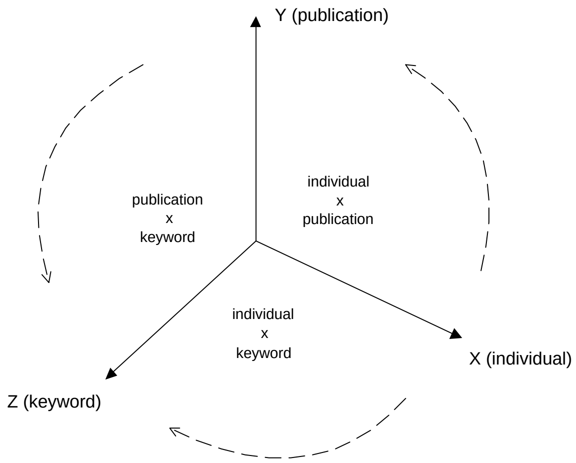

Figure 5. A multidimensional basis for building lattices

Actually, the lattices conceptualise the relation between two of the variables, (x, y), or (x, z), or (y, z). A pair, say (x_0, y_0), will exist in the lattice defined on $X \times Y$ if and only if the following relation composition holds: $R_2(x, z) \circ R_3(z, y)$, meaning that for the pair (x, y), there exists $z \in Z$ such that the composition $R_2 \circ R_3$ holds.

More work remains to be done for investigating this research aspect, linking relations, conceptual space and lattices.

4.2.2 *First question: "Which documents are written about a common topic?"*

In our team first we were interested which documents are about a common topic, as it would be a serious help for our members to know which documents they should consult in their research area within the team. It is also useful for finding similar works that one is not aware of.

In general, we are looking for an interaction between y and z, for example "which documents (y) are on a common topic (z)?". To answer the question we have studied the relations between y and z (see Table 2), and for this we have constructed a Galois lattice.

	ont.	sw	cbr	assoc. rules	bioinfo.	adapt.
cadot03b				x		
cherfi03c				x	x	
daquin02a						x
daquin03a	x	x	x			
lieber02a			x			x

Table 2. Publications $\times$ keywords input table

In this example we have taken some of our publications, but to keep the example in moderated size we have only taken some keywords into account. In the corresponding Hasse diagram (Figure 6) the extension set of concepts contains documents, the intension set of concepts consists of keywords describing these documents. From the diagram we can read the answer to our question by examining the concepts: we can see that {*cadot03b, cherfi03c*} are on *association rules*, {*daquin02a, lieber02a*} are on *adaptation*, and {*daquin03a, lieber02a*} are both written about *case-based reasoning*. In our experiment the constructed lattice for 147 publications and 89 keywords has 253 concepts.

It is possible to extract association rules from the lattice, for example: "bioinfo." $\Rightarrow$ "association rules", which means: every paper that has the keyword "bioinfo." also contains "association rules".

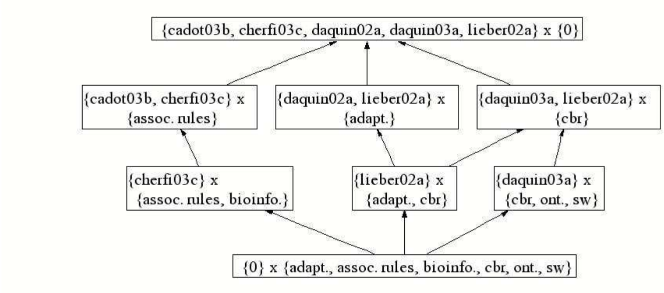

Figure 6. Documents on a common topic

4.2.3 *Second question: "Who are the persons working on a common topic?"*

In the next step we were interested in finding relations between people. For the members in a team it is useful to know who else work on the same topic, whom they can contact to consult on a problem. In this case we need to look for interactions between people and topics (see Table 3). We have constructed another Galois lattice, which differs from the previous example in the sense that the extension set of concepts contains authors. The intension set consists of keywords that are common in publications of all the people present in the extension set.

	repr.	bioinfo.	classification	cbr
Lieber, J.	x	x	x	x
Brachais, S.	x			
Le Ber, F.			x	x
d'Aquin, M.	x	x		x

Table 3. Authors $\times$ keywords input table

Again, the example is voluntarily simplified. From the Hasse diagram (Figure 7) we can read the answers we are interested in, i.e. {*Brachais, S., Lieber, J., d'Aquin, M.*} work on *representation*; {*Le Ber, F., Lieber, J., d'Aquin, M.*} work on *case-based reasoning*; {*Lieber, J., d'Aquin, M.*} work on {*bioinformatics, case-based reasoning, representation*}; and {*Le Ber, F., Lieber, J.*} work on {*case-based reasoning, classification*}. In our experiment we took all the authors and co-authors of our publications (96) and the reduced keyword set (89 keywords). The resulting Hasse diagram has 287 concepts.

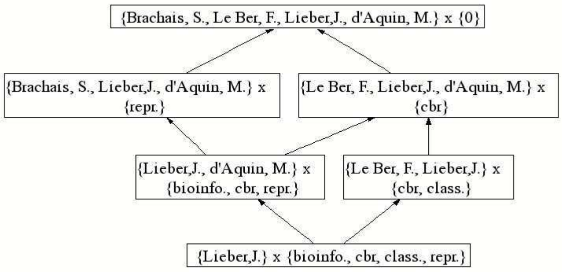

Figure 7. People working on a common topic

4.3 State of the system

For the implementation we have used Java Servlet/JSP technology. Under Java there are several ontology APIs available for RDF(S), OWL, e.g. the Jena2 ontology API [15]. To create ontologies we used the Protégé-2000 editor, and as an ontology language we have used RDF(S). RDF is a datamodel for relations between resources on the Web, and RDF Schema adds vocabulary for RDF. The latest ontology language is OWL, that is an extension of RDFS. It is likely that we will change to OWL soon since it allows better reasoning.

To visualise lattices we have used the freely available Graphviz[6] package that permits to draw the whole graph, but the navigation in a large lattice with more hundred concepts can become difficult. We plan to try other visualisation methods like the fisheye view technique (from the focus node the other nodes are displayed in decreasing levels of detail and at increasing graphical distance), and hypertrees [18, 19].

5 Synthesis and related work

The main problem investigated in our work is the organisation of documents to propose different views enabling a user to discover, to navigate and to query a given corpus. The structure is based on Galois lattices to detect correlations (ex.: authors & subjects). A domain ontology is used to enrich/enhance the organisation process, e.g. filtering the keywords describing a document. We proposed a method of joining an ontology with Galois lattices for concept formation.

For our work we consider three projects as reference works. Galicia is a tool for lattice construction and visualisation. Galicia implements several algorithms in both batch and incremental mode, including iceberg lattices [6]. The Weka tool is a collection of machine learning algorithms for data mining. Weka supports several tasks of the KDD process, including classification and extracting association rules [27]. Both of these tools are implemented in Java and freely available. SEAL (SEmantic PortAL) is a framework to build community web sites. It uses ontologies as key elements for managing community web sites and web portals. SEAL serves as the core methodology underlying the OntoWeb[7] portal. We have chosen these projects for our reference works because our final goal is to create a semantic portal for the team with integrated data-mining capabilities. Building an intranet portal is today a standard first step in knowledge management, and with a portal it is possible to draw together on the desktop all the important information from both inside and outside an enterprise [20].

Pernelle *et al.* developed a system (ZooM) that can give a general view of concepts addressing a large data set. The user selects two ordered nodes in the lattice and ZooM constructs a fine-grained lattice between the antecedents of these nodes [23, 24, 25].

Hotho *et al.* describes a method first clustering documents, then using Formal Concept Analysis on the clusters. Clustering reduces the number of objects such that FCA becomes more efficient [14].

6 Conclusion and perspectives

In this paper we have investigated the application of Galois lattices in a small enterprise, namely within our research team, for guiding the knowledge organisation process. We have proposed to use domain ontologies in several steps of the KDD process. We have shown how to use ontologies for data cleaning to avoid some problems that can arise by an exact keyword-based search. Furthermore, we have used the same ontology to connect it with Galois lattices. We were interested in finding relations between people, documents, and topics in our team.

Our approach can also be used on different data sources to reveal relations between object sets and attribute sets. In the first step a domain ontology has to be built (from scratch or rather reusing already existing ones) that can function as a filter for cleaning the attribute set. In the second step the reduced attribute set can be used in Galois lattices.

We have mentioned that navigation in a large concept lattice can be difficult. There are means to reduce the size of a Galois lattice, for example using iceberg concept lattices. Iceberg concept lattices show only the top-most part of a concept lattice, and they can be used in KDD as conceptual clustering tool, as a visualisation method, as a base of association rules, and as a visualisation tool for association rules [26].

In the next step of our research we plan to investigate the construction of different Galois lattices using our domain ontology. By descending in the ontology from the Top concept and increasing the granularity we will be able to build a family of lattices that would reveal more and more information. We also want to investigate the automatic rule extraction from Galois lattices such as "In 68% of cases, author A has published with author B."

While our keyword ontology is appropriate for our bibliographical items, it is not adequate enough for searching on the Web since the keyword list is not (and cannot be) complete for this task. On the one hand a more complete list has to be built, for example by reusing similar ontologies with an ontology mapping method [16]. On the other hand, some kind of approximate matching may be needed.

We started to build a portal for our team that can only handle internal sources (bibliographical items) at the moment. We want to extend it with the capability to use external sources too to be able to answer questions like "What are the conferences in the next half year in which I may be interested?". The realization of such a system raises several questions but the main idea here is that machines would be searching the Web to find important information. Our view

[6] `http://www.graphviz.org`

[7] `http://www.ontoweb.org`

relies on the Semantic Web principles: the idea of having data defined and linked in a way that it can be used by machines not just for display purposes, but for automation, integration and reuse of data across various applications, and reasoning on documents [2].

REFERENCES

[1] R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. I. Verkamo, 'Fast discovery of association rules', in *Advances in knowledge discovery and data mining*, 307–328, American Association for Artificial Intelligence, (1996).

[2] T. Berners-Lee, J. Hendler, and O. Lassila, 'The Semantic Web', *Scientific American*, **284**(5), 34–43, (May 2001).

[3] C. Carpineto and G. Romano, 'Order-Theoretical Ranking', *Journal of the American Society for Information Science*, **51**(7), 587–601, (May 2000).

[4] Dublin Core Metadata Initiative, 2001. `http://dublincore.org/`.

[5] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, 'From data mining to knowledge discovery in databases', *AI Magazine*, **17**, 37–54, (1996).

[6] GaLicia – Galois lattice interactive constructor. `http://www.iro.umontreal.ca/~galicia/`.

[7] B. Ganter, 'Attribute exploration with background knowledge', *Theoretical Computer Science*, **217**(2), 215–233, (1999).

[8] B. Ganter and R. Wille, 'Applied lattice theory: formal concept analysis', in *General lattice theory*, ed., G. Grätzer, 591–605, Springer Verlag, 2 edn., (1998).

[9] B. Ganter and R. Wille, *Formal concept analysis: mathematical foundations*, Springer, Berlin/Heidelberg, 1999.

[10] R. Godin, R. Missaoui, and H. Alaoui, 'Incremental Concept Formation Algorithms Based on Galois (Concept) Lattices', *Computational Intelligence*, **11**(2), 246–267, (1995).

[11] M. Goebel and L. Gruenwald, 'A survey of data mining and knowledge discovery software tools', *SIGKDD Explorations Newsletter*, **1**(1), 20–33, (1999).

[12] T. R. Gruber, 'A Translation Approach to Portable Ontologies', *Knowledge Acquisition*, **5**(2), 199–220, (1993).

[13] A. Guénoche and I. Van Mechelen, 'Galois Approach to the Induction of Concepts', in *Categories and Concepts: Theoretical Views and Inductive Data Analysis*, eds., I. Van Mechelen, J. Hampton, R. S. Michalski, and P. Theuns, Academic Press Ltd, (1993).

[14] A. Hotho, S. Staab, and G. Stumme, 'Explaining text clustering results using semantic structures', in *Principles of Data Mining and Knowledge Discovery (PKDD 2003)*, pp. 217–228. Springer, (2003).

[15] Jena – A Semantic Web Framework for Java. `http://jena.sourceforge.net/`.

[16] Y. Kalfoglou and M. Schorlemmer, 'Ontology mapping: the state of the art', *The Knowledge Engineering Review*, **18**(1), 1–31, (2003).

[17] M. Klemettinen, H. Mannila, P. Ronkainen, H. Toivonen, and A. I. Verkamo, 'Finding interesting rules from large sets of discovered association rules', in *Proceedings of the third international conference on Information and knowledge management*, pp. 401–407. ACM Press, (1994).

[18] J. Lamping and R. Rao, 'Laying out and visualizing large trees using a hyperbolic space', in *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pp. 13–14. ACM Press, (1994).

[19] J. Lamping, R. Rao, and P. Pirolli, 'A focus+context technique based on hyperbolic geometry for visualizing large hierarchies', in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 401–408. ACM Press/Addison-Wesley Publishing Co., (1995).

[20] A. Maedche, S. Staab, N. Stojanovic, R. Studer, and Y. Sure, 'SEmantic PortAL - The SEAL Approach', in *Spinning the Semantic Web*, eds., D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster, MIT Press, Cambridge, MA, (2002).

[21] D. L. McGuinness, 'Ontologies Come of Age', in *Spinning the Semantic Web*, eds., D. Fensel, J. Hendler, H. Lieberman, and W. Wahlster, MIT Press, (2002).

[22] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, 'Efficient mining of association rules using closed itemset lattices', *Inf. Syst.*, **24**(1), 25–46, (1999).

[23] N. Pernelle, MC. Rousset, H. Soldano, and V. Ventos, 'Zoom: a nested galois lattices-based system for conceptual clustering', *Journal of Experimental and Theoritical Artificial Intelligence (JETAI)*, **14**(2), 157–187, (september 2002).

[24] N. Pernelle, MC. Rousset, and V. Ventos, 'Construction et affinement de hiérarchies à partir de données semi-structurées', *Revue d'extraction des connaissances et apprentissage*, **1**(1-2), 81–92, (2001).

[25] N. Pernelle, V. Ventos, and H. Soldano, 'Zoom: Alpha galois lattices for conceptual clustering', in *Managing Specialization/Generalization Hierarchies (MASPEGHI)*, (October 2003).

[26] G. Stumme, R. Taouil, Y. Bastide, N. Pasquier, and L. Lakhal, 'Computing Iceberg Concept Lattices with TITANIC', *Data and Knowledge Engineering*, **42**(2), 189–222, (2002).

[27] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufmann Publishers, 1999.